

Citation for published version:

Johnson, S & Rasulova, S 2017, 'Qualitative research and the evaluation of development impact: incorporating authenticity into the assessment of rigour', *Journal of Development Effectiveness*, vol. 9, no. 2, pp. 263-276.
<https://doi.org/10.1080/19439342.2017.1306577>

DOI:

[10.1080/19439342.2017.1306577](https://doi.org/10.1080/19439342.2017.1306577)

Publication date:

2017

Document Version

Peer reviewed version

[Link to publication](#)

This is an Accepted Manuscript of an article published by Taylor & Francis in *Journal of Development Effectiveness* on 12 April 2017, available online: <http://www.tandfonline.com/10.1080/19439342.2017.1306577>

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Qualitative research and the evaluation of development impact: incorporating authenticity into the assessment of rigour

**Susan Johnson, Centre for Development Studies, University of Bath
Saltanat Rasulova, Oxford Policy Management**

Abstract

Renewed attention is being paid to the quality of qualitative research in studies of development impact in the wake of intense debate over the role of randomised control trials. These debates raised the bar over quality while also provoking concerns about whose voices matter and the politics of evidence-based policy making. This paper argues that both of these issues can be addressed through Guba and Lincoln's trustworthiness criteria, and particularly the principle of authenticity which was developed to respond to demands for research to be transformative and emancipatory. Adopting these criteria in commissioned evaluations as checklist 'artefacts' are a means to improve rigour and raise ethical standards.

Corresponding author:

Susan Johnson, Dept of Social and Policy Sciences, University of Bath, Bath, BA2 7AY
Email: s.z.johnson@bath.ac.uk

Co-author: Saltanat Rasulova, Oxford Policy Management, [Level 3, Clarendon House](#)
[52 Cornmarket Street, Oxford, OX1 3HJ](#)

Key words: qualitative research, rigour, trustworthiness, impact evaluation, evidence-based policy

Words: 7922

Acknowledgement: This paper is one of the outputs of a two-year collaborative research project investigating ways to improve rigour of qualitative research in impact evaluation between the University of Bath and Oxford Policy Management. Saltanat Rasulova was a Research Associate at the University of Bath from 2014-2016 when the main research for this paper was undertaken. This project operated through a Knowledge Transfer Partnership (KTP009285) funded jointly by Innovate UK and the Economic and Social Research Council. We would like to acknowledge this support as well as comments on earlier versions from Laura Camfield, James Copestake and Ian MacAuslan.

1. Introduction

Renewed attention is now being paid to the role of qualitative methods in evaluations of the impact of development policies and interventions. This arises in the wake of a decade of intense debate in the development sector over the “experimentalist surge” (Picciotto 2014) of randomised control trials (RCTs). These have been used in a wide range of sectoral interventions ranging from education to agriculture, financial services to political empowerment, with results claimed to be the basis for evidence-based policy making. Their claim to be the ‘gold standard’ of rigour has raised the bar the use of qualitative methods in impact research. The white heat of this debate is now waning amidst claims that RCTs are “off the gold standard” and with calls for the debate to be refocussed on the core functions and purposes of impact evaluation and for the deployment of “relevant rigour” which is argued to require negotiation (Camfield and Duvendack 2014, Guijt and Roche 2014). Moreover, the narrow view of RCTs as evidence for policy making, has provoked an important wider concern about its politics. The question of how such - apparently value-free - evidence arises, is critical when development is in fact conceptualised as a transformational process based on rights, with politics and power relations at its centre and hence a concern for means and not only ends (Eyben and Guijt 2015).

Practically also, recognition that a huge range of development interventions and policies are not evaluable using experimental techniques, has led to moves to broaden the range of methods (Stern, Stame et al. 2012) to incorporate state-of-the-art approaches involving theory-based and realist methods, along with ways of making qualitative work rigorous for causal inference (White 2009, Stern 2015). These approaches make use of a range of both quantitative and qualitative evidence but rely very extensively on qualitative methods. This leads to a broader concern regarding the rigour with which research using qualitative data collection and analysis is undertaken (Spencer, et al. 2003).

This paper therefore sets out to consider how rigour in qualitative impact evaluation can be better addressed. According to the Oxford Living English Dictionary, rigour is defined as “the quality of being extremely thorough and careful” (OED online, 2017). Guidelines for evaluating the quality of qualitative evaluation have been developed for various contexts such as Spencer et al. (2003) for evidence being reviewed in relation to UK government policy development, and more recently Santiago-Delefosse (2016) specifically focussed on the health sector. The latter highlights the importance for quality of the core steps of any research exercise: starting with clarity over theoretical frameworks; research questions; aims and objectives; literature review; methodology and design; sampling; data; analytical process; and ethics. It indicates some consensus on criteria that are more specifically

associated with discussion of qualitative research quality: reflexivity, credibility and transferability.

These latter criteria of credibility and transferability draw from the language and ideas of Guba and Lincoln's early framework for assessing quality (1981; 1982) which they termed trustworthiness. This offered to the naturalist paradigm a parallel approach to the positivist criteria for rigour. However, this remained unsatisfactory as it remained within the "third generation" of evaluation based in the judgement of external actors. They therefore proposed "fourth generation" evaluation underpinned by the principle of authenticity and embedded in the constructivist paradigm (1989). This principle recognises that all judgement and understanding has a value basis and hence is subject to political and ideological forces. The consequence for evaluation is that it must involve negotiation among stakeholders so that the basis of their understandings and values is part of the process. While their initial position was that third generation evaluation based in a positivist perspective could not be mixed (2001) with such fourth generation approaches, their later position (2007, p25) recognises that the authenticity principle and its concerns for fairness and negotiation were helping to address ethical and ideological problems which "while at first appearing to be radical...are nevertheless becoming mainstream". However, it is notable that the checklists for qualitative research discussed above do not incorporate this principle and therefore remain firmly within the domain of third-generation judgement so neglecting an overt concern with the politics involved.

This paper is particularly concerned with the extensive body of qualitative research which is commissioned by funders to evaluate and establish the impact of their funded policies and interventions. In the context of this renewed focus on rigour - or reliability - in qualitative research for impact assessment, this paper returns to examine the debate over rigour and the trustworthiness criteria of Guba and Lincoln as core principles. We then review the authenticity principle in greater depth to engage with the politics of evaluation by bringing in the voices of beneficiaries and stakeholders. While the underlying thinking is familiar within debates over participatory approaches to appraisal and evaluation, and more specifically in participatory action research, the scope for explicit participatory approaches in commissioned evaluations is usually very limited or - more specifically - not perceived as what the commissioner may be seeking. Indeed these evaluations fit the definition of what Macdonald called "bureaucratic" evaluation (Norris 2015). We therefore specifically focus on the space for practicing authenticity within such contexts and argue that the time is ripe for actively adopting authenticity into checklists of rigour as an evaluation "artefact" which can be used for "playing the game to change the rules" (Guijt 2015).

The paper starts by reviewing the underlying basis of concerns about rigour in qualitative research, leading to different views of whether and how it can be addressed. It introduces

Guba and Lincoln's principles of trustworthiness and then focuses on the transformative-
emancipatory values that the principle of authenticity engages with. It then discusses the
scope for practicing these in the organisational context of commissioned research. The
paper concludes by discussing the ethical challenges and how the principle might be
promoted in the development sector by its use as an evaluation "artefact" (Guijt 2015).

2. Qualitative research and the 'problem' of rigour

The philosophy underlying most qualitative research is that there is no single truth or reality
and that phenomena depend on our perceptions and interpretations of them. Therefore the
nature of reality is not unique or objectively verifiable but relative (ontologically relativist)
and is created by our interpretations of it (epistemologically constructionist / interpretivist /
or subjective). The "truth" presented is then a result of the interaction and relationship
between the researcher and the researched rather than simply of the research design, as it
is constructed by individuals under particular conditions and in a particular context, and
consequently cannot be generalised (Sandelowski, 1986). This contrasts with quantitative
research in which there is an objective reality (ontological realism) driven by natural laws
which science seeks to discover. In this case for investigation - epistemology - the researcher
and the object of inquiry are separate (duality) therefore the researcher does not influence
the way this truth is experienced. Truth is therefore seen as being obtained through
scientific methods in which theories are proposed, and hypotheses derived from them are
tested, with the potential for their falsification (Teddie and Tashakkori, 2009) so presenting
findings or knowledge as generalisable and relevant across time and context.

Although this is a standard way of defining the contrast between qualitative and quantitative
methodologies, the field of qualitative research is diverse and expanding with a range of
approaches and methods such that it involves further debate and disagreement. Traditions
range from phenomenology, ethnography, case study, biography, grounded theory, to
feminist and Marxist perspectives and so on, in turn having diverse research aims.
Moreover, how qualitative researchers apply constructionism and/or interpretivism depends
on what traditions they belong to, and some qualitative traditions can be closer to positivism
and post-positivism (Racher and Robinson, 2002) in understanding the world as having
objective reality, while others may differ between the extent to which they seek to
understand or to bring *about a social change* (Sandelowski, 1986). So a straightforward
dichotomous understanding of quantitative as being premised on positivist/post-positivist
paradigm and qualitative on interpretivism is misleading.

With this plurality, and despite the widespread use of qualitative research and contribution
to understanding, it has faced significant criticism. In the field of impact evaluation, this

criticism was particularly made by Archie Cochrane (1972) now seen as the 'father' of randomized control trials as the 'gold standard' for evidence. He saw qualitative research as unscientific and unfit for purpose. The very features of it that are its strength such as the depth and closeness of the researcher to the research topic and to the data are argued to be its weakness due to their being subjective and lacking rigour (Morse et al, 2002).

Nevertheless, it is the frequently profound submersion of the researcher with the data collection from research participants and context which are its advantages. As Pretty (1995, p.40) suggests 'we can only get a human idea of what is in the world, and so science itself can only be a human picture of the world' and according to Rolfe (2002, p. 91) truth is when it 'rings true', that is when it resonates with our own experiences'. Despite this contribution, and the counter-arguments by qualitative researchers who point out that quantitative research suffers from subjectivism too when it comes to choosing categories and variables (Flyvbjerg, 2011), it is qualitative research that has been put on the back foot from the perspective of wider policy discussion and the debate over evidence.

The response from qualitative researchers has been a debate over whether and how rigour can in fact be implemented and the question of whether rigour can or should actually be a criterion is itself highly contested. While some take the view that neglecting rigour could undermine the whole existence of this paradigm as a systematic science (Morse et al., 2002); others argue that it does not need to be discussed at all (Sparkes, 2001). Moreover, given the range of qualitative methodologies there is little agreement as to whether a set of criteria needs to be general to all the fields of qualitative research or distinctive to each one of them (Sandelowski, 1993).

The debate over criteria involves three positions (Denzin 2011). First, the foundationalists – who take the view that all research can conform to shared criteria (Teddle and Tashakkori, 2003); second, the quasi-foundationalists who argue that a specific set of criteria unique to qualitative research is needed (eg. Sandelowski, 1993; Lincoln and Guba, 1985); and finally the non-foundationalists who argue that it is understanding rather than prediction that is necessary and that 'inquiry and its evidence is always political and moral' (Denzin, 2011) and therefore that criteria are not appropriate.

For those who consider a set of criteria as useful, one starting point is the quality criteria used in quantitative research. The central ones are internal validity, objectivity, reliability and generalization (Bryman, 1988). Internal validity refers to the suitability of a chosen method to measure what it is used to measure i.e. the degree to which the results are attributable to the independent variable; objectivity is achieved by positioning the researcher outside of the measured activity; reliability is when the same results can be

achieved using the same methods; generalisation is ensured through using statistical tests to control other factors (Hamberg et al., 1994).

A range of quality principles have been developed (Agar, 1986; Guba, 1981; Kirk and Miller, 1986; Leininger, 1985; Brink, 1991) from which it is evident that there is considerable overlap in the nature and language of the criteria being presented as well as some mapping onto the quantitative criteria. Agar's (1986) credibility, accuracy of representation, and authority of the writer are versions of validity and reliability and are close in meaning to the two typologies of Guba (1981) and Lincoln and Guba (1985). Guba's (1981) model of four aspects of trustworthiness that are relevant to both quantitative and qualitative studies are very similar to the one by Lincoln and Guba (1985) except for the principle of authenticity. An exception is Kirk and Miller (1986) whose criteria argue for consistency of results and stability over time which seems contrary to a qualitative view that interpretations and understandings may change over time.

In the literature in this field it is Lincoln and Guba's (1985) criteria that have become the most used and cited with their language of "trustworthiness" rather than rigour also being adopted (see table 2). The comparative advantages of this typology are seen as its relevance across qualitative paradigms (Sale and Brazil, 2004) and having the advantage of parsimony (Bryman et al., 2008). The typology is the most developed and therefore we use it for the conceptual and analytical clarity and depth that it can offer.

3. Trustworthiness principles

Table 1 summarises the criteria and we discuss the first four which have been most used in practice and turn to the authenticity principle in the next section.

Credibility is argued to be similar to internal validity in quantitative research. Research is seen as credible when the researcher has confidence in the truth of the findings with regard to the subjects of research and the context where it was conducted. As there is understood to be a plurality of truths within the qualitative paradigm, then credibility is about 'representing those multiple realities revealed by informants as adequately as possible' and testing those realities against various groups of people (Krefting, 1991; p. 215), so as to draw out common themes, actions and issues. The description of realities is important, so that people who share similar experiences should recognise them and be able to relate them to their own realities (Sandelowski, 1986). Moreover, Sandelowski (1986) suggests that the credibility of qualitative work increases when the researcher discusses his/her own behaviour alongside the behaviour of other 'subjects of research' and reflects on their closeness to the data in order to recognise how she might be unduly giving weight to particular positions or interpretations over others when analysing and interpreting it. So the

researcher needs to stay close to data but at a reflexive distance from it. This needs to result in such an interpretation of data which is recognisable for both the research participants and scientific community.

Table 1. Lincoln and Guba's principles for evaluating trustworthiness in qualitative research

Qualitative research	Questions that underpin the principles of qualitative research (Pretty, 1994; p. 42)	Quantitative research concepts
Credibility	How can we be confident about the 'truth' of the findings?	Internal validity
Confirmability	How can we be certain that the findings have been determined by the subjects and contexts of the inquiry, rather than the biases, motivations and perspectives of the investigator?	Objectivity
Dependability	Would the findings be repeated if the inquiry were replicated with the same (or similar) subjects in the same or similar context?	Reliability
Transferability	Can we apply these findings to other context or with other groups of people?	Generalisation
Authenticity	Have people been changed by the process? To what extent did the investigation prompt action?	

Source: own table based on Pretty (1994) and Lincoln and Guba (1985)

Confirmability is about ensuring that the research process and findings are not biased, hence it refers to both the researcher and the interpretations (Baxter and Eyles, 1997). This raises the issue of distance from and influence of the researcher on data collection and analysis when the researcher is actively involved with research participants and constantly engages with the data. This closeness of the researcher to the object of the study is argued to be a unique feature of qualitative data, so it is challenging for the researcher as a 'positioned subject' (Rosaldo, 1989) to consciously reflect about her own acts and background in relation to the data. Confirmability is achieved when the interpretation of data is neutral and free from the researcher's personal bias.

Dependability is to ensure consistent data collection without unnecessary variations to ensure repeatability of the research process. This is about being able to trace sources that the data comes from and about documenting the data, methods and decisions made during the fieldwork. So consistency in the entire research process is key for achieving

dependability. This criterion is suggested to be closely linked to credibility and be equally important for qualitative research. Lincoln and Guba (1985) themselves argue that establishing dependability ensures credibility. Although the strategies used to prove these two principles can be similar, their meanings are different: credibility is about 'accurate representation of experience while dependability focuses attention on the researcher-as-instrument and the degree to which interpretation is made in a consistent manner.' (Baxter and Eyles, 1997; p. 517).

Transferability is similar to external validity or generalisability but is not reached through random sampling and probabilistic reasoning. Transferability is when the research descriptions and findings are sufficient to draw similarities with another context. To achieve this criterion, the researcher needs to provide detailed descriptive information. These details should enable the reader to judge the applicability of findings to her own settings. However, as discussed, it is hard to generalise meanings constructed from a small group of respondents and this is usually not the goal of qualitative research. Nevertheless, the researcher has the (ethical) responsibility to describe the findings in the way that allows transferability and let the reader decide whether those meanings are transferrable to her context (Baxter and Eyles, 1997).

Despite the acknowledged advantages of these criteria, it is unsurprising given the paradigmatic disagreements laid out above, that there are criticisms. The first is that in order to argue for absolute trustworthiness it is necessary to take an epistemologically positivistic view (Pretty, 1994) rather than that truth is valid for a certain time and a certain context and is not static. Therefore, in order to ensure trustworthiness it is suggested that the criteria should include the process of inquiry, informing the reader about what has and has not happened before, during and after the inquiry (ibid).

In addition, Morse et al. (2002) argue that confirmability is not suitable for phenomenology, feminist research and critical theory 'in which the investigator's experience becomes part of data, and which perceive reality as dynamic and changing' (p. 19). Smith (1993 in Tobin and Begley, 2004) argues that some procedures such as member or dependability checks (in which researchers share their conclusions with respondents and get feedback on their conclusions) are not appropriate to the philosophical idea of multiple realities that qualitative research is based on and therefore, target its fundamental epistemological and ontological premise.

As indicated above, the non-foundationalists reject Lincoln and Guba's criteria out of hand. As Barbour (2001) and Morse et al. (2002) claim, any technical solutions reducing qualitative research to a list of methodological procedures does not ensure rigour. Barbour (2001) argues that such procedures can only be successful if they are conducted within a proper

understanding of qualitative research as creative and flexible, which makes it problematic to reduce it to what is seen as a box-ticking exercise. Moreover, it is argued that the procedural aspects must come second when evaluating the quality of qualitative research after 'moral principles and ethical standards' of how researchers relate to research participants (Rallis et al., 2007). This ethical argument is paramount since qualitative research is based on relationships which emerge between the researcher and data, the researcher and research participants, data and research participants and the wider circle of readers. Therefore, they argue that the trustworthiness of a qualitative study needs to be judged on how ethically it is done with relation to research participants, other stakeholders and the scientific community.

4. The authenticity principle

It was in response to these criticisms that Lincoln and Guba added the fifth criterion of authenticity in order to more clearly distinguish these principles from positivistic ones (Lincoln, 1990 cited in Pretty, 1994). The authenticity criterion has five elements (Guba and Lincoln, 1989) and was developed in accordance with a constructivist tradition i.e. knowledge is relative to time and place, subjective meanings matter and that '...truth is a matter of consensus among informed and sophisticated constructors...' (Lay and Papadopoulos, 2007; p. 495).

Schwandt et al (2007) argue that authenticity is an extension of the trustworthiness criteria because it enables questions to be asked about how interpretations are made and how this process has evolved. Indeed, the authenticity principle recognises that inquiry and understanding are a process of learning, changing, negotiating and ultimately acting. Qualitative research affects the consciousness of the researcher and research participants to the extent that it can change the way they understand the truth(s). For the authenticity principle, the evaluation research should recognise and promote the diversity of the value systems and people's constructions of the world which change constantly as people interact with one another. Therefore the principle takes into account the process of forming interpretations from the value-point of respondents, their voice, their diversity, their positioning and empowerment towards other respondents and themselves. In this process, research participants as well as researchers learn to respect each other, to see the issues from different perspectives and consequently, negotiate the construction of truth. Based on a constructivist epistemology, the main drive for the authenticity principle is negotiation i.e. negotiation of constructs and values. Guba and Lincoln (2001) suggest that such an approach to the evaluation can deal with the conventional problems with evaluations which, among other weaknesses, favours the point of view of the funder, which is disempowering, unfair to other stakeholders, and fails to accommodate value-pluralism, and is overcommitted to the positivist paradigm of inquiry.

Table 2 gives the summary of the five criteria of authenticity including their short definitions. The assumptions behind the authenticity principle are that people and groups have different value systems which affect their constructions and the five dimensions seek to expose the ways in which these have been sought out and developed through the research process. First is fairness which they say involves a core dimension of balance. This means making sure that different constructions emerge, and so allowing conflicting constructions and value structures to be expressed. This therefore means that the researcher must actively observe and seek these out and that this then leads into open and active negotiation around the agenda for action.

The second two dimensions consider the ways in which participant's and stakeholder's 'conscious experiencing of the world' (Guba and Lincoln cited in Schwandt, Lincoln, and Guba 2007, p. 22) develops and their constructions change as they gain experience and interact with others. The second dimension of ontological authenticity captures the way that the experience affects research participant's own understanding of the phenomenon under scrutiny. This arises as the process of enquiry itself exposes their own views and understandings in ways that offer new perspectives and can be captured through their own testimonies of these changes. The third dimension captures the idea that it is not only their own understanding that changes but also their appreciation of how other's view the world: what Guba and Lincoln term educative authenticity. So allowing greater appreciation of different opinions, judgments, and actions.

The fourth dimension of catalytic authenticity is defined as the extent to which action is stimulated and facilitated by the evaluation. It captures the idea that the change in understanding of one's own and other's perspectives has the potential to lead to new action: actions which derive from the potential for negotiation over tensions and contradictions. They comment that this is usually lacking at the level of participants and stakeholders, rather it being centralised through processes of consolidation and dissemination.

Finally, the criterion of tactical authenticity evaluates the extent to which individuals are in fact empowered to take the action that this new understanding involves and in relation to the final negotiated agenda for action.

For all these components of the authenticity principle to be achieved, negotiation is essential and it is achieving consensus that drives the principle. The principles therefore respond to the concerns of the non-foundationalists and critical theorists that research in itself has political and moral dynamics whose purpose may be to affect the world and not

simply to represent it in objective ways. They affect the world by changing understanding with the potential to lead to action.

In the context of the assessment of the impact of development interventions this principle therefore makes central the issue of the rights of those affected by a policy or intervention to be engaged in the research and to engage with other policy stakeholders in negotiation over findings and implications. Indeed, as Guba (1987; p. 39) suggests authentic evaluation will *'...essentially be about the process of negotiation with and among stakeholders, and that the product of evaluation is not '...a series of conclusions and recommendations, but as an agenda for further negotiation'* (emphasis in original). The principle therefore particularly responds to the recent concerns raised regarding the politics of evidence in development and fears that the dominance of experimentalist approaches has become increasingly hegemonic in the practices and processes of the allocation of aid resources (Eyben and Guijt 2015). We therefore explore further the potential for its use in the context of commissioned evaluations.

Table 2. The authenticity principle and strategies to achieve it

Criteria	Definition (Guba and Lincoln, 1989)	Strategies to achieve and evidence authenticity
Fairness	The extent to which all competing constructions have been accessed, exposed, and taken into account in the evaluation report, that is, in the negotiated emergent construction.	Observing contradictions and tensions among constructions of stakeholders
		Negotiating these contradictions and tensions and establishing an 'agenda for subsequent action' (Guba & Lincoln, 1989, p. 246).
Ontological Authenticity (world around)	The extent to which individual's constructions (including those of the evaluator) have become more informed and sophisticated	Analyzing statements (e.g., testimonies) provided by the research participants and leaving audit trails that document the participant's growth in consciousness and understanding of the world, as well as the growth in the researcher's own 'progressive subjectivity' (p. 248).
		Observing the participants in action for collecting evidence of an elevation in the participant's consciousness level.
Educative authenticity (other people's constructions)	The extent to which individuals (including the evaluator) have become more understanding (even if not more tolerant) of the constructions of others.	Analyzing statements (e.g., testimonies) provided by the research participants and leaving audit trails that document the participant's growth in understanding of and appreciation for [but not necessarily agreement of] the constructions of <i>others</i> outside their stakeholding group are enhanced" Guba & Lincoln, 1989, p. 248, italics in original).
Catalytic authenticity	The extent to which action (clarifying the focus at issue, moving to eliminate or ameliorate problems, sharpening values) is stimulated and facilitated by the evaluation.	Obtaining testimony from all participants and stakeholders regarding their interest in and willingness to turn their increased understanding into action.
		Obtaining testimony regarding the joint actions of participants who have come to resolutions stemming from negotiations of tensions invoked by contesting and contradictory constructions of the stakeholders' (Onwuegbuzie, Leech, & Collins, 2008, p. 9).
		Assessing the extent to which the actions that stemmed from the increased understandings that emerged during the course of the study actually occurred
Tactical authenticity	The extent to which individuals are empowered to take the action that the evaluation implies or proposes	Obtaining testimony from all the participants and stakeholders regarding whether/ how the emergent feelings of empowerment evolved and manifested themselves
		Systematically following up within a predetermined time frame to assess which participants and/or groups acted on their increased feelings of empowerment, and what actions came to the fore
		Researcher and participants (jointly) assess the degree of empowerment that evolved during the study.

Source: Authors' own table based on Guba and Lincoln, 1989; Onwuegbuzie, Leech, & Collins, 2008. .

5. Practicing authenticity

The authenticity principle gives considerable weight to the ability of evaluation research to incorporate the values and constructions of participants as well as empowering them to improve their situations. Such intentions are of course not without problems when it comes to practice. In the development field there has been a long discussion of participation in both development practice (Cooke and Kothari 2001, Hickey and Mohan 2004) and research (Chambers 2004) which have highlighted the core problems of structural power dynamics in attempts to involve participants in empowering ways. In the context of commissioned evaluations, evaluators are usually employed for their independence and even where they are able to involve a wide range of local stakeholders, often operate in the field with virtually no involvement of the commissioners themselves. Hence, the space for affecting commissioners own constructions of the reality of the intervention is often extremely limited or indeed non-existent.

Additionally the practical context of most of commissioned evaluations stand largely outside a participatory paradigm that would most effectively respond to these concerns. There are real practical challenges regarding the amount of time spent in the field together with research participants and the ability to develop effective relationships with them such that these objectives can be achieved. While one-off interviews or group discussions and workshops for disseminating research findings can engage new thinking and help inform, they clearly fall short of this vision and rarely focus on participants and stakeholders becoming enlightened about each other's values and constructions. This makes it less likely that the research would be able to generate far reaching impact on research participants to fulfil the authenticity principle. More realistic would be to assume that it is rather the researchers or evaluators who would be more likely to experience changes in understanding during the course of the study. Moreover, it is hard to foresee that any change would be sustainable enough to lead to catalysed action.

While the authenticity principle promotes recognition of the diversity of values and constructions, it is unclear how exactly consensus is to be reached or how disagreements are dealt with. There will always be power relations among different research groups of participants and stakeholders.. Indeed, qualitative research in international development evaluation rarely seeks to act as a solution provider, consensus builder or producer of tangible action. It is issues such as these that approaches to evaluation ranging from empowerment (Fetterman and Wandersman 2012) to participatory (Coghlan and King 2005) and democratic evaluation traditions (House and Howe 2000, MacDonald and Kushner 2005) endeavour to manage through specific ways of conducting them.

However, with this somewhat pessimistic assessment, the point here is to ask how far these aspects can be taken in the context of commissioned assessments that are constrained by bureaucratic requirements? The point is to find ways that consultants undertaking commissioned evaluations can exercise their agency within these processes to raise the bar.

At the moment, the strategies suggested in Table 2 represent a range of rather ideal approaches for achieving such ethically complex ideas as participation and empowerment. In this sense, while the authenticity principle speaks to an important normative concern that research be undertaken in a transformative-emancipatory way, the principle needs to bring into view the wider research context and practical organisational framework for implementing its strategies. It is these wider dynamics that have been at the core of recent concern about the dominance of RCTs in the evidence-based movement and how it meets visions of development based in rights-based approaches and the transformation of power relations (Eyben and Guijt 2015).

Indeed, engaging with an evaluation assessment in order to open up the space for debate and learning is potentially dangerous terrain for participants and raises the possibility of extremely unsatisfactory or even dangerous processes of engagement to result. This may be, at a minimum, by wasting the time of participants in processes whose potential for producing change is virtually non-existent or extremely marginal. Moreover, in the context of power relations surrounding the implementation of policies or programmes, the result may be even more problematic with the potential for negative consequences if findings lead to decisions and action that detrimentally rebounds on them. This could be the withdrawal of resources or worse, the potential for victimisation if particular actors feel aggrieved at information revealed during evaluation processes. These issues raise in a specific way the questions of what have been called “macro-ethical issues” which go beyond the “micro-ethical issues” of consent, anonymity and so on. This concern is systemic towards how evaluation relates to society and social processes more broadly; whose interests it serves; its role in deliberation and the use of evidence and in public sector management and wider public debate (Barnett and Munslow 2014):13). Indeed it also raises the question of how development is done and the contribution of evaluation to ethical development (Barnett and Munslow 2014). While in principle therefore, authenticity incorporates some aspects of concern around the politics and power dynamics of impact evaluation, strategies for implementation are obviously much more difficult. This discussion therefore highlights the need for a wider ethical framework within which the authenticity principle must operate at two levels: the individual study and commissioning organisation itself, and, second, the wider institutional context of evaluation practice.

For commissioned researchers, guidelines and procedural measures cannot fully mould their actions and they are largely driven by their own values and ethics in the context of

operational practical and structural constraints. Ethics are particularly important as researchers and evaluators (especially commissioners) tend to have resources and power concentrated in their hands, such that the scope for negotiation that can take place in the course of a study, is likely to be extremely constrained by its design from the outset. Therefore, the issue of ethics is one that extends well beyond procedural ethical requirements such as consent. From the view of the authenticity principle it means that the researcher must be fair in allowing a diversity of constructions and values to be expressed, she must actively seek to become more informed of her own as well as other people's constructions through reflexivity, and encourage the right actions towards the phenomenon by generating trustworthy and ethical qualitative findings.

Going further to create space for the practice of negotiation as part of the practical strategies can obviously be of value under the right circumstances. It can legitimate space for diversity of views within the evaluation in the context of problematic top-down power relations and in principle requires space for negotiation with commissioners themselves. How then can this be done?

As Guijt (2015) argues, it is necessary to “play the game to change the rules” in part by using the “artefacts” of evaluation themselves. That is, the processes, mechanisms and tools that are deployed in this field. It is necessary to understand them, “reframe” and “intelligently adopt them” (p200). We therefore propose that commissioned researchers deploy checklists for rigour in this way. That they include the authenticity principle as a means of raising the game with commissioners as to what is appropriate and ethical in the conduct of evaluation research and seek to make more space for engagement, negotiation and action. While we recognise that a list of strategies is not ideal if it is solely treated in the spirit of box ticking – including them can create space for those researchers and commissioned organisations who do indeed wish to add value and undertake qualitative research for assessing impact in a more progressive way. So that while it is not possible to up-end power relations at a single turn, it is certainly the case that demonstrating a concern for these ethical dimensions and their practice can help confirm and give space for reflexivity over the intention of the exercise. This in turn calls for the practices of commissioning organisations to respond in terms of their own guidelines and checklists.

Table 3. A principle of authenticity and initial strategies to achieve it

Principle	Stages of implementation	Strategies to achieve the principle	Purposes of strategies
Authenticity (negotiation over	Design	Purposeful sampling	To have a wide ‘representation’ of people who have ‘relevant’ relationship with the phenomenon

constructions through diversity of views)		Enabling space for voice and negotiation of potential directions of the research	To include the views of research participants on the research design e.g. to identify hard to access respondents, to be responsive to the local mechanisms of consensus building and negotiation
	Data collection and analysis	Making space for a wide group of research participants to participate in data collection and analysis	To draw out diverse views and experiences of the phenomenon among research participants and negotiate the articulation of findings with them
	Dissemination and follow up	Offering means through which research participants can be engaged with commissioners as stakeholders in making recommendations and taking follow up actions	To make sure that research participants have their 'say' in recommending any changes to the intervention and taking part in their follow up. Any recommendations should be responsive to the diversity of values and opinions of research participants
	Review	Re-visiting commissioners, stakeholders and respondents to assess what has changed.	To establish how the evaluation has been used and examine what changes in understanding and action were achieved by all parties involved.

As a starting point then the orientation is to find ways (see Table 3) for: (i) making space for voice and negotiation by research participants as stakeholders within the design and conduct of the research through meetings and workshops and ensuring these involve excellent facilitation skills; (ii) finding ways to make space for participation in data collection and analysis; (iii) offer means through which representatives of these views can be engaged with commissioners in later stages of the evaluation. Indeed, as an aspiration, getting commissioners themselves actively involved in more transparent negotiations with participants and other stakeholders over action and outcomes would be a key achievement.

All of these steps must involve ethical considerations regarding the potential for harm. In this context, the minimum standard is always to have considered the issues against the checklist and explained why the approach taken is appropriate within the resources available and how it avoids potential harm. Table 3 presents a starting checklist for orienting research to use the authenticity principle.

The final point in Table 3, addresses the question of what ultimately changed both in terms of understanding of the issues and the actions of commissioners and stakeholders. This

effectively means a further review. In many programme areas this is in fact possible since there are frequently follow on programmes whose planning stages could deploy a review approach, but of course it is rarely done. Commissioners usually start again and do not wish to revisit past programmes or projects, especially if these were not perceived to be 'successful'.

Of course checklists themselves guarantee very little and to achieve change they have to be underpinned by a particular mind-set and ethics (Barbour 2001). This mind-set is an orientation underpinned by a concern for ethical practice at both the "micro" and "macro" levels oriented towards inclusion, negotiation and deliberation. It is this orientation that underpins the question of "how much?" of any of these strategies is appropriate. There is always space for more to be done and it is constant reflexivity on these issues that is required which in development involves constant concern for the dynamics of power relations (Groves and Hinton, 2004).

Moreover, organizational culture is also vital for the effective use of checklists. Indeed it is the process of arriving at a checklist within a discussion of practice that is most likely to bring about change in organisational - and indeed industry wide - practice (Bosk, Dixon-Woods et al. 2009). Nevertheless, the adoption of checklists does offer a minimum standard which is to have considered the issues against the checklist and explained why the approach taken is appropriate within the resources available and for the avoidance of potential harm.

More broadly, checklists can act to raise the issues for debate. For consultants, the use of checklists of rigour, and indication of how they are to be applied, can be used to raise the understanding of quality in their tenders. Moving beyond this, the principle of authenticity demands the creation of space for learning among all stakeholders, for interaction and negotiation. This means experimenting with approaches to expand the scope for authenticity and engaging with commissioners over the processes and resources involved. We argue that the time is ripe for actively adopting it into checklists of rigour as an evaluation "artefact" as a step towards "play(ing) the game to change the rules" (Guijt 2015).

Conclusion

This paper has been motivated by the recent call for qualitative research used to assess the impact of development interventions to address the concern for rigour in the context of debates about the politics of evidence. The paper has therefore reviewed the literature on rigour in qualitative research, starting with a review of the underlying philosophical basis to

explain the disagreements within the debate between qualitative methodologists themselves over the feasibility of rigour.

Notwithstanding these problems, considerable efforts have been made to develop standardised criteria for evaluating rigour and the context of impact evaluation research is sufficiently practical to require them. This paper returned to Guba and Lincoln's framework for trustworthiness for the reason that it is the most developed and used and is appropriate to the context of much impact assessment research - particularly that in development consultancy - which operates within a post-positivist paradigm of realist ontology and modified objectivist epistemology where qualitative methodologies and evidence are still assessed through criteria of representation and objectivity.

We focused in more detail on the often over-looked principle of authenticity. Authenticity is embedded in concerns of the constructivist camp for research that is transformative-emancipatory in orientation, and fits well with concerns in evaluation about the politics of evidence and the wider macro-ethical concerns as to the role of evaluation in society more broadly. In order to move forward practically we propose that adding authenticity to checklists of rigour is therefore a strategic and useful way to deploy the "artefacts" of evaluation protocols and guidelines to demonstrate approaches to rigour to commissioners. At the same time there is the potential to use this approach to challenge the commissioners and institutional frameworks surrounding evaluation itself.

Given the concern that checklists themselves are inadequate, we argue that using such a principle requires an orientation to better practice with these ethical concerns at their source. It is this orientation that underpins the question of "how much?" of any of these strategies is appropriate. There is always space for more to be done and it is constant reflexivity on these issues that is required. The minimum standard is always to have considered the issues against the checklist and explained why the approach taken is appropriate within the resources available and the avoidance of potential harm. Moving beyond this, the principle of authenticity demands the creation of space for learning among all stakeholders, for interaction and negotiation. This means experimenting with approaches to expand the scope for authenticity and engaging with commissioners over the processes and resources required.

References:

- Agar, M. (1986). *Speaking of ethnography*. Beverly Hills, CA: Sage.
- Barbour, R. (2001) Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? *BMJ*. 2001;322: 1115–1117
- Barnett, C., & Munslow, T. (2014). Workshop Report: Framing Ethics in Impact Evaluation: Where are we? Which route should we take? *Evidence Report No 98*. Institute for Development Studies, Sussex www.ids.ac.uk.
- Baxter, J. and J. Eyles, (1997) Evaluating qualitative research in social geography: establishing 'rigour' in interview analysis. *Transactions of the Institute of British Geographers*, 22(4): 505–525.
- Bosk, C. L., Dixon-Woods, M., Goeschel, C. A., & Pronovost, P. J. (2009). Reality check for checklists. *The Lancet*, 374(9688), 444–445. doi: 10.1016/S0140-6736(09)61440-9
- Brink, P. (1991) Issues of reliability and validity. In: Morse, J. (ed) *Qualitative nursing research: a contemporary dialogue*. Sage, London, pp. 164–186
- Bryman, A. (1988) *Quantity and Quality in Social Research*, London: Unwin Hyman
- Bryman, A. Becker, S. and J. Sempik (2008) Quality Criteria for Quantitative, Qualitative and Mixed Methods Research: A View from Social Policy' *Int. J. Social Research Methodology*, 11 (4): 261–276
- Camfield, L., & Duvendack, M. (2014). Impact Evaluation - Are We 'Off the Gold Standard'? [Article]. *European Journal of Development Research*, 26(1), 1–11. doi: 10.1057/ejdr.2013.42
- Chambers, R. (2004). Ideas for development: reflecting forwards. IDS Working Paper 238.
- Coghlan, A. T., & King, J. A. (2005). Participatory Evaluation. In S. Mathison (Ed.), *Encyclopedia of Evaluation* (pp. 292–297). Thousand Oaks, CA: Sage Publications, Inc.
- Cooke, B., & Kothari, U. (eds) (2001). *Participation: the new tyranny?* London: Zed Press.
- Cochrane A. L. (1972) Effectiveness and Efficiency: Random Reflections on Health Services, The Nuffield Provincial Hospitals Trust. (Reprinted in 1989 in association with the BMJ, Reprinted in 1999 for Nuffield Trust by. London: The Royal Society of Medicine Press; 1972.
- Denzin, Norman K. (2011) "The Politics of Evidence." In *The Sage Handbook of Qualitative Research*, edited by Norman K. Denzin and Yvonna S. Lincoln, 645–58. London: SAGE, 2011.
- Eyben, R., & Guijt, I. (2015). Introduction. In R. Eyben, I. Guijt, C. Roche & C. Shutt (Eds.),

The politics of evidence and results in international development : playing the game to change the rules? Rugby: Rugby : Practical Action Publishing.

Fetterman, D. M., & Wandersman, A. (2012). *Empowerment Evaluation Principles in Practice*: Guilford Publications.

Flyvbjerg, B. (2011) Case study. In Denzin, K. N. and Lincoln, S. Y. (eds) *The sage handbook of qualitative research*. USA: Sage Publications. pp. 301-316

Groves, L. C., & Hinton, R. B. (2004). *Inclusive aid changing power and relationships in international development*. London: Earthscan.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology Journal*, 29 (2), 75-91.

Guba, E. G., and Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco, CA: Jossey-Bass.

Guba, E. G., and Lincoln, Y. S. (1982). Epistemological and methodological bases of naturalistic inquiry. *Educational Communication and Technology Journal* 30 (4): 233-252

Guba, E. G. and Lincoln, Y. S. (1989) *Fourth Generation Evaluation*, Newbury Park, CA: Sage Publications.

Guba, E.G. and Lincoln, Y.S. (1994) Competing paradigms in qualitative research. In. Denzin, N.K and Y.S. Lincoln (eds.), *Handbook of qualitative research*. pp. 105-117

Hamberg K, Johansson E, Lindgren G and Westman G. (1994) Scientific rigour in qualitative research—examples from a study of women's health in family practice. *Family Practice* 11: 176-181.

Hickey, Samuel, and Giles Mohan. *Participation: From Tyranny to Transformation? Exploring New Approaches to Participation in Development*. edited by Samuel Hickey and Giles Mohan London: London : Zed, 2004.

House, E. R., & Howe, K. R. (2000). Deliberative democratic evaluation. *New Directions for Evaluation*, 2000(85), 3-12. doi: 10.1002/ev.1157

Kirk, J. and Miller, M. (1986) *Reliability and validity in qualitative research*. Beverly Hills, CA: Sage.

Krefting, L. (1991) Rigor in qualitative research: The assessment of trustworthiness. *American Journal of Occupational Therapy*, 45: 214-222.

Leininger, M. M. (1985). Nature, rationale and importance of qualitative research methods in nursing. In Leininger, M. M. (ed), *Qualitative research methods in nursing*. New York: Grune & Stratton, pp. 1-28

Lincoln, Y. (2005). Institutional review boards and methodological conservatism. The challenge to and from phenomenological paradigms. In Denzin, N. and Y. Lincoln (eds.), *The Sage handbook of qualitative research* Thousand Oaks, CA: SAGE. 3rd ed., pp. 165-181

Lincoln, Y. and E. Guba (1985) *Naturalistic Inquiry*. Beverly Hills: Sage,

MacDonald, B., & Kushner, S. (2005). Democratic Evaluation. In S. Mathison (Ed.), *Encyclopedia of Evaluation* (pp. 110-114). Thousand Oaks, CA: Sage Publications, Inc.

Morse J.M., Barrett M., Mayan M., Olson K. and Spiers J. (2002) Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods* 1(2), Article 2.

Morse J.M. (1999) Myth 93: Reliability and validity are not relevant to qualitative inquiry. *Qualitative Inquiry* 9: 717–718

Norris, N. (2015). Democratic evaluation: The work and ideas of Barry MacDonald (Vol. 21, pp. 135-142).

Onwuegbuzie, A. J., Leech N.L. and Collins K.M.Y., (2008). "Interviewing the Interpretive Researcher: A Method for Addressing the Crises of Representation, Legitimation, and Praxis." *International Journal of Qualitative Methods* 7(4): 1-17

Picciotto, Robert. "Is Impact Evaluation Evaluation?". *European Journal of Development Research* 26, no. 1 (2014): 31-38.

Pretty, J. N. (1994) Alternative systems of inquiry for sustainable agriculture. *IDS Bulletin*, 25(2): 37-48. IDS: University of Sussex

Racher F. and Robinson S. (2002) Are phenomenology and postpositivism strange bedfellows? *Western Journal of Nursing Research* 25: 464–481

Rallis, S., Rossman, B and Gajda, R. (2007) Trustworthiness in evaluation practice: An emphasis on the relational. *Evaluation and Program Planning* 30: 404–409

Rolfe G (2002) 'A Lie that Helps us See the Truth': Research, truth and fiction in the helping professions. *Reflect Pract* 3:89–100.

Rosaldo, R. 1993 (1989). Grief and a headhunter's rage. In: *Culture and truth*. London & New York, Routledge: 1-21.

Sale, J. and K. Brazil (2004) A Strategy to Identify Critical Appraisal Criteria for Primary Mixed-Method Studies *Quality & Quantity* 38: 351–36

Sandelowski M. (1993) Rigor or rigor mortis: the problem of rigor in qualitative research revisited. *Advances in Nursing Science*, 16:1–8.

Sandelowski, M. (1986) Problem of rigor in qualitative research. *Advancing in Nursing Science* 8(3): 27-37.

Santiago-Delefosse, M., et al. (2016). "Quality of qualitative research in the health sciences: Analysis of the common criteria present in 58 assessment guidelines by expert users." *Social Science & Medicine* 148: 142-151.

Schwandt, T. A., Lincoln, Y. S., & Guba, E. G. (2007). Judging interpretations: But is it rigorous? trustworthiness and authenticity in naturalistic evaluation. *New Directions for Evaluation*, 2007(114), 11-25. doi: 10.1002/ev.223

Sparkes, C. (2001). Myth 94: Qualitative health researchers will agree about validity. *Qualitative Health Research*, 11(4): 538-552.

Spencer, L., Ritchie, J., Lewis, J., & Dillon, L. (2003). Quality in Qualitative Evaluation: a framework for assessing research evidence: UK Cabinet Office.

Stern, Elliot, Nicoletta Stame, John Mayne, Kim Forss, Rick Davies, and Barbara Befani. "Broadening the Range of Designs and Methods for Impact Evaluations " In *Report of a study commissioned by the Department for International Development* London: DFID Working Paper No. 38, 2012.

Teddie C. and A. Tashakkori (2009) *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Thousand Oaks, CA: SAGE

Tobin, G and Begley C (2004) Methodological rigour within a qualitative framework. *J Adv Nurs* 48(4): 388–96

White, H., & Phillips, D. (2012). Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework: 3ie <http://www.3ieimpact.org/>.